

Citation for published version:

Pan, S, Giannikas, V, Han, Y, Qiao, B & Grover-Silva, E 2017, 'Using Customer-related Data to Enhance E-grocery Home Delivery', *Industrial Management and Data Systems*, vol. 117, no. 9, pp. 1917-1933.
<https://doi.org/10.1108/IMDS-10-2016-0432>

DOI:

[10.1108/IMDS-10-2016-0432](https://doi.org/10.1108/IMDS-10-2016-0432)

Publication date:

2017

Document Version

Peer reviewed version

[Link to publication](https://doi.org/10.1108/IMDS-10-2016-0432)

The final publication is available at Emerald via <https://doi.org/10.1108/IMDS-10-2016-0432>

University of Bath

Alternative formats

If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Using Customer-related Data to Enhance E-grocery Home Delivery

Purpose – The development of e-grocery allows people to purchase food online and benefit from home delivery service. Nevertheless, a high rate of failed deliveries due to the customer's absence causes significant loss of logistics efficiency, especially for perishable food. This paper proposes an innovative approach to use customer-related data to optimize e-grocery home delivery. The approach estimates the absence probability of a customer by mining electricity consumption data, in order to improve the success rate of delivery and optimize transportation.

Design/methodology/approach – The methodological approach consists of two stages: a data mining stage that estimates absence probabilities, and an optimization stage to optimize transportation.

Findings – Computational experiments reveal that the proposed approach could reduce the total travel distance by 3% to 20%, and theoretically increase the success rate of first-round delivery approximately by 18%-26%.

Research limitations/implications – The proposed approach combines two attractive research streams on data mining and transportation planning to provide a solution for e-commerce logistics.

Practical implications – This study gives an insight to e-grocery retailers and carriers on how to use customer-related data to improve home delivery effectiveness and efficiency.

Social implications – The proposed approach can be used to reduce environmental footprint generated by freight distribution in a city, and to improve customers' experience on online shopping.

Originality/value – Being an experimental study, this work demonstrates the effectiveness of data-driven innovative solutions to e-grocery home delivery problem. The paper provides also a methodological approach to this line of research.

Paper classification – Research paper

Keywords – City Logistics, Food Delivery, E-commerce, Data Mining, Freight Transportation.

1 Introduction

Recent developments of e-commerce have had a significant impact on food supply chains. Today, many traditional grocery retailers offer their customers the opportunity to purchase food items online and have them delivered to their home utilizing their existing distribution network (Ogawara et al., 2003, Agatz et al., 2008). At the same time, new companies enter the retail groceries market by providing online supermarkets with no physical stores, fulfilling home deliveries from their warehouses (e.g. the case of *Ocado* (Saskia et al., 2016)). In addition to that, internet-based retailers, like Amazon, exploit their e-commerce expertise to build their own online grocery shops thus extending even more the options end-customers have for purchasing groceries online (Kang et al., 2016).

In e-grocery commerce, home delivery –the process of delivering goods from a retailer’s storage point (e.g. distribution centers, shops) to a customer’s home– plays a crucial role (Punakivi and Saranen, 2001). In fact, due to its convenience to customers, home delivery has become a dominant distribution channel of business-to-consumer e-commerce (Campbell and Savelsbergh, 2006). Nevertheless, a certain challenge faced in e-grocery is that the perishability and storage condition-sensitivity of food and drink items requires the attendance of the customer at the moment of delivery (Hsu et al., 2007). At the same time, this makes alternative methods for unattended delivery, such as delivery boxes, reception boxes and shared reception boxes, hard and unsafe to use. This has led e-tailers to introduce strict policies for deliveries that could not be completed due to the customer’s non-attendance (Ehmke, 2012a), while aiming to increase the probability of an attended delivery by allowing their customers to choose their preferred time slot. However, it is still common for end-customers to be absent at the time of delivery either due to their own fault (e.g. failing to remember) or due to a delayed delivery (e.g. due to traffic).

In this paper, we aim to address the attended home delivery problem (AHDP) (Ehmke, 2012a, Ehmke and Campbell, 2014) in e-grocery motivated by the fact the attendance of the customer is often hard to predict. We do this by investigating a new approach that utilizes customer-related data to improve attended home delivery efficiency. The approach consists of two stages. The first stage concerns a data

mining process, whose objective is to estimate the purchaser's absence probability at a given time window according to his/her electricity consumption behavior. The second stage uses the calculated absence probabilities as an input to an optimization model for managing the fleet of trucks that execute the home deliveries.

This paper aims to make both a theoretical and a practical contribution to the AHDP. With regards to the theoretical contribution, this study is among the first ones applying data mining techniques to AHDP. It provides a novel methodology to investigate the AHDP from the aspect of customer-related data, which can be thought of as a new research line on AHDP. From a practical point of view, the two-stage approach proposed could serve as a decision-making model for e-grocery retailers, or other retail businesses that provide (attended) home delivery service, to organize or enhance their delivery service.

Following this introduction, Section 2 provides a review of related work. Then, Section 3 presents the two-stage approach, that is, respectively, data mining stage and the transportation planning stage. Section 4 presents an application example to demonstrate the practicability and performance of the proposed approach. Finally Section 5 concludes this work.

2 A brief review of e-grocery and its logistics

In this section we briefly discuss the literature from three related problems: current challenges to grocery and food e-commerce, delivering grocery and food items in e-commerce, and innovations in home delivery.

2.1 On-line grocery retailing

On-line grocery retailing, also known as e-grocery, is a type of business to consumer e-commerce that has enjoyed great growth in the last decade and is expected to continue growing in the years to come (Mortimer et al., 2016). Similarly to other Internet retailing examples, on-line grocery shopping offers significant benefits to end-customers including time-savings, access to multiple retailers and products and home delivery. Nevertheless, there are a number of factors that can affect the decision of end-customers to use on-line channels for their grocery shopping. These important factors include:

- (1) *Ordering interface and product information* (Tanskanen et al., 2002, Boyer and Hult, 2005, Boyer and Hult, 2006): a well-designed and easy to use online shop is critical to the overall customer experience of online shopping. In grocery shopping in particular, where customers are used to visually checking the products they buy (for nutrition details, expiry dates, ingredients etc.), it is important that label information and accurate product photos are available to the end-customer via a website or app. Moreover, customers often expect that managing their basket and checking out online should be a straightforward process, similar to placing products in a shopping cart and visiting a cashier in a physical store.
- (2) *Product range availability* (Colla and Lapoule, 2012, Anckar et al., 2002, Zott et al., 2000): customers expect the products they can normally purchase in a physical store to be available online. A certain limitation here has to do with the purchase of non-packaged food (or even fresh packaged food) products, especially in cases when the customer is used to buying a product after visually checking its condition.
- (3) *Suitable logistical delivery options* (Wang et al., 2016a, Anu et al., 2012, Koster, 2003): having chosen the products they want to purchase, customers need to also make a decision about the delivery of the physical goods. Besides choosing the preferable option, this factor also includes the retailer's response to delivery problems, cost of different options and subscription models for delivery services.
- (4) *Consistency between all sales and media channels* (Ishfaq et al., 2016, Hübner et al., 2016, Breugelmans and Campo, 2016): this is a relatively recent trend due to the emergence of omni-channel retail. With retailers offering more and more options for customers to purchase goods, they need to make sure that different channels offer the same information and functionality.

In this paper, we focus on the third factor described above – the delivery options offered to the customer – due to its continuous importance for e-commerce success (Ricker and Kalakota, 1999, Lee and Whang, 2001, Zhang et al., 2011, Ramanathan, 2010). We discuss this challenge in more detail in the next section.

2.2 Delivery challenges in grocery and food e-commerce

Unlike traditional in-store sales where customers are able to receive physical products directly after their purchase, e-grocery requires a set of logistics operations that are crucial not only for the right delivery of a product but also for the overall satisfaction of the end-customer (Hübner et al., 2016). It has been noted that the retailers that provide a grocery home delivery service are the ones that face the greatest logistical challenges (Fernie et al., 2010). These logistical challenges refer to both the back-end fulfillment and the last mile distribution of an order (Hübner et al., 2016). Back-end fulfillment mainly deals with the picking and preparation of an order. Last mile distribution (which is the focus of this study) involves decisions related to the delivery method, time and area as well as the returns of unwanted products.

Two methods are commonly available for e-grocery delivery: *home delivery* and *click and collect*. As regards the first method, due to the perishability and storage condition-sensitivity of food, attendance of customer (or receiver in general) is often required at the moment of home delivery (Hsu et al., 2007). This problem is known in the literature as the Attended Home Delivery Problem (AHDP) (Ehmke, 2012a, Ehmke and Campbell, 2014). As the attendance of a customer is hard to predict, home delivery usually results in high rate of failures (Agatz et al., 2011, Gevaers et al., 2011, Lowe and Rigby, 2014) and can lead to high delivery costs, waste of waiting time (for the customer) and waste of energy spent in transportation. Effectively and efficiently tackling AHDP is becoming a key success factor to food delivery, as well as an important challenge with regards to sustainability of freight transportation (Ehmke and Mattfeld, 2012, De Marco et al., 2014, Gevaers et al., 2011).

At the stage of transportation, the AHDP has mainly been addressed as a Vehicle Routing Problem (VRP), or VRP with time windows (VRPTW) if a delivery time window is imposed (Ehmke and Mattfeld, 2012, Hsu et al., 2007). The latter is also studied as time slot management problem (Agatz et al., 2011). Reliability and width of the time windows – slots – strongly impact on the results of routing optimization as well as on customer's experience. At the stage of reception, some practical solutions can be employed when a customer is absent at delivery (i.e. unattended delivery), such as leaving the package to someone nearby (neighbors, gatekeeper etc.) or at a secure place (mailbox, yard, garage etc.) or calling the purchaser to confirm

attendance (Punakivi and Saranen, 2001, Ehmke, 2012a). These solutions are helpful to the problem, but their disadvantages are significant; time wasted for phone calls or for waiting for delivery, product security issues, lockers storage condition issue, reception boxes size issue (Iwan et al., 2016, Lowe and Rigby, 2014, Ehmke and Campbell, 2014). These disadvantages are particularly noticeable in food delivery.

In order to tackle the above challenges and improve home delivery, e-tailers and logistics providers have been driven to invest in delivery innovation and technology (Lowe and Rigby, 2014). In the competitive delivery market, this is expected to create solutions that can better meet customer needs. We discuss relevant innovations in the next section.

2.3 Innovation in home delivery

Innovation in home delivery has received considerable attention in both real-world practice and research in recent years. Generally, innovations can be classified in three lines of research that are constantly and interactively developing: organizational, technology-enabled, and data technique-enabled innovations. Organizational innovations refer to the implementation of innovative organization models or methods for last mile distribution (LMD). Examples include urban consolidation centers (Allen et al., 2014, Van Duin et al., 2010), synchronization and horizontal collaboration (de Souza et al., 2014) and crowdsourcing (Paloheimo et al., 2015, Chen et al., 2016). Technology-enabled innovations are the result of the application of emerging technologies in LMD. Automated lockers and drones are the most noticeable examples in this line as both industrial practice (DHL, 2016, Amazon, 2016) and academic work indicate (Gevaers et al., 2011, Iwan et al., 2016).

Data technique-enabled innovations refer to applications of data techniques (data mining, data analytics, big data etc.) in LMD aiming at improving effectiveness and efficiency. Academic research has concentrated heavily on the exploitation of historical traffic data sets to optimize time-dependent routes in LMD. For example, Ehmke and Mattfeld (2012) and Ehmke et al. (2012) propose mining taxi-Floating Car Data (FCD) to determine time-dependent travel times in a city and use that information to optimize routing. Chen et al. (2016) mines taxi-FCD to plan taxi-based

crowdsourcing LMD in big cities. These examples have indicated a great potential for data techniques to improve LMD, especially in the context of city logistics. Nevertheless, data other than historical traffic data are rarely studied, even though they could provide significant benefits to LMD. We notice here that customer-related data have been previously used in the context of e-commerce but mainly for the purpose of marketing. Examples include mining customer data to understand consumption performance and design online shopping services (Liao et al., 2011), or for the purpose of customer relationship management (Mithas et al., 2005, Karakostas et al., 2005).

Inspired by the potential data-enabled techniques have shown for LMD in particular but also for logistics and supply chain management in general (Wang et al., 2016b), this study is focused on this type of home delivery innovation. More specifically, the fact that the attended home delivery could benefit from predictions regarding a customer's absence makes data-enabled techniques - which are capable of providing such predictions - a promising tool for tackling this problem.

3 A data-driven approach for e-grocery home-delivery

The discussion in the previous section has highlighted the importance of the AHDP problem in e-grocery delivery as well as the potential of data usage in last-mile logistics. Motivated by this, this work introduces an innovative approach that utilizes customer-related data to improve the home-delivery service of grocery items. In summary, we attempt to exploit customer-related data in order to provide a method to determine the optimal time slot profiles for carriers or shippers, in order to optimize the success rate of home delivery.

The approach proposed in this paper consists of two stages (see Figure 1). The first stage aims to estimate the probability the purchaser of an online grocery order will be at the chosen delivery location at different points of time during the day. At this stage, customer-related data are collected and used as an input to a data mining model, capable of calculation the customer attendance (or absence) probabilities. A time series of customer telemetry data is taken as the input to the framework. The time series is arranged as a temporally ordered telemetry measurement collected from each customer, noted as $(x_1, x_2, x_3...x_n)$, as x_i to be the measurement obtained at the i -th time

slot during the whole observation time window. The physical sense of x_i depends on the intrinsic of different telemetry sources. For example, if we use in-house electrical power loads data of each customer, x_i then denotes the instant power load level of one particular customer measured at the i -th time slot. Other examples include water consumption data and historical data from previous deliveries. With more and more data being collected nowadays from households and individuals (due to emerging trends like the Internet of Things, Big Data and Open Data), we expect the availability of customer-related data that can be used in this approach to increase considerably in the years to come. Such an open dataset on electricity consumption is provided by the Irish Commission for Energy Regulation (CER) and is used as an example in the remaining of this study.

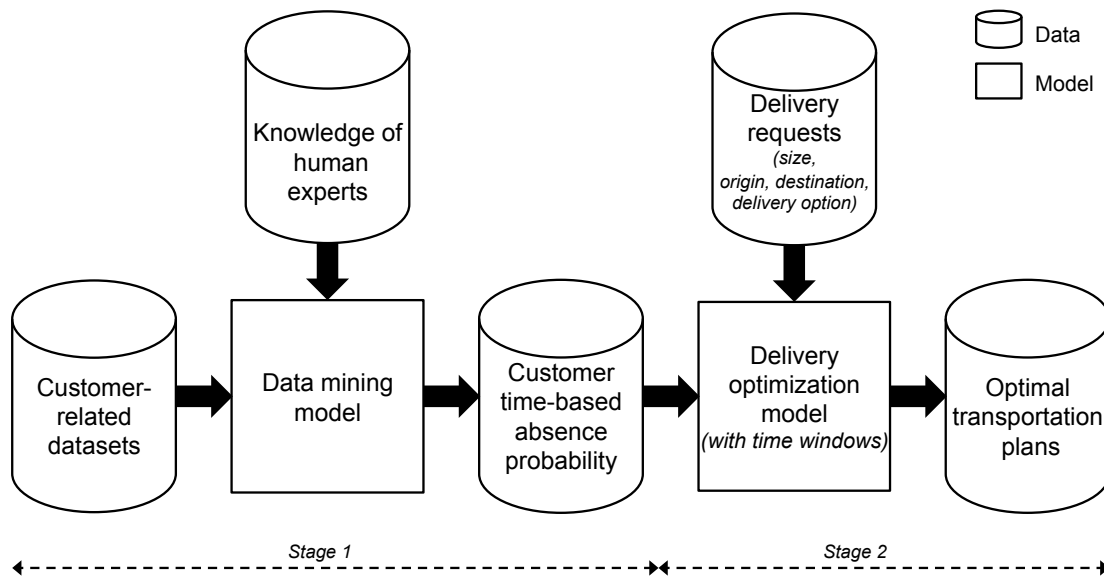


Figure 1. Methodology Flowchart

Once the customer-related data are collected, they are used in a data mining model to estimate home attendance. As an extra input, expert knowledge might also be required at this stage, especially with regards to energy consumption, in order to link consumption data with occupancy detection (home attendance or absence). In this paper, we estimate home absence at given time windows according to one's electricity consumption behavior. Building statistical correlations between energy consumption traces and social-economical factors of customers has received a lot of attention lately. The major purpose of the research topic is to find out the underlying patterns of customers' daily living habits based on their energy usage behaviors, in order to

provide a strong base to not only the energy supplying service, but also to other value added services, such as on-line shopping recommendation and targeted advertisement.

Generally, the previous research efforts in this domain may be divided into two categories: end-user models and econometric methods. End-user models are commonly used as an alternative to black-box methods (Wood and Newborough, 2003, Abreu and Pereira, 2012, Aman et al., 2011, Kolter and Ferreira Jr, 2011, Beckel et al., 2012). They require information about housing conditions, electrical appliance usage and environmental factors. Such background information is used with energy domain knowledge to disaggregate the daily electricity consumption measurements of a specific user into elementary components, including heating/cooling, water usage, cooking and other behaviors. The disaggregation result is then applied to find out their usage preferences. The shortcoming of end-user models is that forecasting performance depends heavily on the quality of available information, which makes them sensitive to noise and unable to perform automatically. Econometric methods (Beckel et al., 2012, Abreu and Pereira, 2012, Kolter and Ferreira Jr, 2011) estimate the relationship between energy consumption profiles and the factors influencing consumption behavior using statistical machine learning approaches, such as support vector regressors, decision trees and so on. Econometric models are built by learning the mapping from pairs of the factors and energy consumption profiles automatically, which is appealing for realistic application deployments. Recently, this category has gained popularity. Most research efforts along this direction focus on estimating the users' general social economical factors, such as professions, family status, salary levels and so on.

The outcome of the first stage, which is in the form of a set of time windows with attendance probabilities for each customer, serves as an input to the second stage. In this stage, the approach aims to provide transportation plans for carriers who need to schedule their home-delivery operations. Here, information about delivery requests, such as origin and destination, delivery option and order size is combined with the estimated attendance probabilities to decide when and how a delivery should be made. As already discussed in Section 2.2, scheduling of attended home deliveries is often done by first modeling the problem as a vehicle routing problem (usually with time windows and limited capacity). In short, in the context of last mile delivery, the

vehicle routing problem involves the planning of a set of routes for a capacitated vehicle that starts from and ends with a (retailer's) warehouse, aiming at delivering the demanded goods to a set of customers (Ehmke, 2012b, Cattaruzza et al., 2015).

Once the VRP problem is modeled, an optimization technique can be employed to solve it. In this paper, we propose a model that takes into account the customer absence probabilities and aims to establish transportation plans that maximize delivery success rates while minimizing total transportation distances. Other models can also be considered in the second stage of the approach, depending on the key performance indicators defined by each retailer or logistics provider as well as the limitations imposed by each business scenario.

Having introduced our approach, in the next section we will demonstrate its application using an example. We will also evaluate the effectiveness of the approach (and thus the effectiveness of data-driven techniques for e-grocery home delivery) by comparing it with a base scenario where absence probabilities are unknown.

4 Application example

This section aims to illustrate how the proposed methodological approach can be applied in a real world case. It also demonstrates potential benefits of the approach over conventional solutions. The illustration of the application follows the two-stage approach, in which Stage 1 relates to the customer-related data collection and mining; and Stage 2 to the transportation planning optimization.

4.1 Stage 1

4.1.1 Customer-related data collection

Due to the scarcity of available customer-related data, this work uses a publicly available electricity consumption trace data set, named CER ISSDA to simulate and demonstrate the use of estimating occupancy patterns of private households for estimating absence probabilities. The CER ISSDA dataset is collected by the Irish Commission for Energy Regulation (CER) in a smart meter study¹: it contains electricity consumption data of 4,225 private households and 485 small and medium enterprises (all called customer hereafter); the trace covers 1.5 years (from July 2009

¹ <http://www.ucd.ie/issda/data/commissionforenergyregulationcer/>

to December 2010). For each customer, the daily load curve is sampled every 30 minutes: energy data can be thought of as a series of timestamps and energy readings. In addition to energy data, the dataset includes a series of survey sheets and answers for each customer, describing their housing condition, occupancy, employment status, income level, social class, appliance usage information and other socio-economic factors. Forty-one survey questions belonging to five categories were selected that enable the customer profiles to be built based on heating and lighting behavior, hot water and other electrical appliances usage. As a sample for this work, we select 20 private households randomly from the dataset to simulate the shipping network in a practical application. To avoid the effects of seasonal variation on the consumption profiles, e.g. air conditioning use in summer and heating use in winter changing daily consumption patterns, we only focus on the time period from April 1st to June 30th in 2009 to collect electricity consumption data, which in total involves 91 days.

4.1.2 Data mining model

The detection of occupancy within a residential household is primarily based on activity detection. Activity within a household is often linked to electric consumption (Kleiminger et al., 2013). Therefore, the variance and magnitude of a residential consumption profile can give indications about whether a household is occupied or not. Peaks and high variation in a consumption profile (indicating typical active behaviors) were labeled in our model and then compared to low variation and low overall consumption periods. The combination of low overall consumption and low consumption profile variation indicates a period of inoccupation (absence). Automated devices within a household can interfere with occupancy detection. High overall consumption due to heating devices left on can also interfere with occupancy detection. Thus it is necessary to identify patterns for each specific client. For each client, a typical profile variation for occupied and unoccupied periods was established in our model. A threshold of average consumption was also determined. When overall consumption was below the threshold and variation of the profile was relatively low, the period was determined to be unoccupied.

Based on the above qualitative analysis, we propose a computational model to estimate occupied periods for a given user, as described below. Hereafter, we use $X_{i,j}$ to denote electricity consumption level at the j -th time step of i -th day for one specific

user. Given the context of shipping programming, we especially focus on estimating occupancy period within the time interval ranging from 8:00 am to 8:00 pm. Our estimation approach is defined into 4 successive steps:

1. Indicator estimation: calculating the consumption magnitudes $X_{i,j}$ and the absolute values of the first-order difference $D^1_{i,j}$ and $D^2_{i,j}$, as defined in Eq (1). The three measurements form an indicator cell $(x_{i,j}, D^1_{i,j}, D^2_{i,j})$, which is used as the feature to detect occupancy patterns. Given all 91 days in the database, we have in total 91 (days) * 12 (delivery hours per day) * 2 (timestamps per hour) indicator cells for each user.

$$D^1_{i,j} = |x_{i,j} - x_{i,j+1}| \text{ and } D^2_{i,j} = |x_{i,j} - x_{i,j-1}| \quad (1)$$

2. Label the L_p indicator cells for each user with the most typical active behaviors manually by human experts on energy consumption. Treating occupancy detection as an anomaly detection problem, the labeled L_p indicator cells form a reference set to identify whether the customer's activities are present within a given time period. Based on the labeled indicator cells, we build a linear kernel one-class SVM based detector F (Schölkopf et al., 2001). The output $Y_{i,j}$ of the SVM detector is 1 or 0, deciding whether the human activity is present at the given time step or not (1 corresponds to the presence of the customer and vice versa).

$$Y_{i,j} = F(x_{i,j}, D^1_{i,j}, D^2_{i,j}) \quad (2)$$

3. Apply the built SVM based detector on the 24 time steps between 8:00 am to 8:00 pm for each day besides the labeled L_p time steps. The binary output of the detector is used as an estimated occupancy label of the concerned time steps. As such, for the i -th day, we can obtain a 24-dimensional binary valued vector as the estimated occupancy label vector of the concerned day.
4. Accumulate estimated occupancy label vectors derived from the total 91 days. For each day of the week (from Monday to Sunday), we calculate the empirical expectation of the occupancy label for each time step as the occupancy probability of the time step j for each day of the week. As a result, we generate a probability map $M \in \mathbb{R}^{24 \times 7}$. Each entry $M_{k,j}$ represents the estimated occupancy probability of the time step $(j+16)$ on the given day of the week. Monday to Sunday is denoted as 0 to 6.

According to the algorithmic description, the occupancy detection procedure is defined as an one-class anomaly detection problem in our work. Compared with varied inactive consumption behaviors, typical active behavior patterns indicating residences' occupancy is easier to identify manually for human experts. Therefore, the expected occupancy detector is built to describe the common characteristics of the active consumption profiles and expected to differentiate the active profiles from inactive ones at the same time. This is a typical one-class learning problem in machine learning theory (Schölkopf et al., 2001) and one-class SVM perfectly matches with this goal. In previous research, it is popularly used to detect a concerned events from backgrounds given only limited number of samples belonging to the event are available. The indicator cell designed in this work covers both the instant measurement of the electricity consumption level and the dynamical information about the consumption variation, represented by the two first-order differences. The feature design is based on the theoretical analysis of active consumption profiles. High instant consumption level and high variance of consumption measurement are strong indicators of active behaviors, thus the chosen three features are able to sensitively indicate the potential human activities existing behind the consumption measurements. Finally, the binary output of the occupancy detector can be noisy due to the hard threshold intrinsic. Furthermore, to insert occupancy estimate as a constraint into the shipping programming problem, we need to smooth the binary decision into a soft, continuously valued confidence of residences' occupancy for each specific time step. As a result, empirical expectation is used as the estimate of the underlying true occupancy probability.

4.1.3 Customer absence probability

Finally, the results of time slots with absence probability by customer are obtained, as Figure 2 shown. Among the 20 customers randomly selected, 5 are excluded in this study because the variation of their consumption curve is not significant (thus making these cases irrelevant to this study, e.g. absence probability always $> 90\%$). Moreover, the daily working hours from 8h-17h (common delivery hours in many countries) are chosen. The customer attendance probability is then the inverse of the absence probability (knowing that both staying out or inactive at home are considered as absence here). The results show that most customers have similar absence probability curve during most days of the week but Saturday. The figure shows only the average

of weekdays' probability to compare with Saturday (Sunday is not considered for delivery).

| | C1 | C2 | C3 | C4 | C5 | C6 | C7 | C8 | C9 | C10 | C11 | C12 | C13 | C14 | C15 |
|---------|-----|-----|-----|-----|-----|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 8H | 67% | 75% | 80% | 35% | 93% | 85% | 81% | 80% | 44% | 51% | 28% | 30% | 56% | 46% | 51% |
| M 8H30 | 41% | 84% | 79% | 38% | 89% | 91% | 78% | 72% | 67% | 37% | 52% | 48% | 58% | 43% | 61% |
| o 9H | 27% | 90% | 75% | 38% | 82% | 91% | 81% | 66% | 77% | 37% | 56% | 49% | 46% | 44% | 65% |
| n 9H30 | 24% | 94% | 75% | 55% | 79% | 92% | 82% | 64% | 84% | 37% | 64% | 52% | 46% | 47% | 65% |
| d 10H | 30% | 89% | 81% | 67% | 80% | 90% | 85% | 67% | 85% | 37% | 68% | 66% | 37% | 63% | 64% |
| a 10H30 | 40% | 90% | 83% | 68% | 77% | 88% | 85% | 66% | 87% | 40% | 72% | 64% | 25% | 67% | 59% |
| y 11H | 57% | 86% | 80% | 73% | 80% | 86% | 80% | 71% | 85% | 45% | 75% | 66% | 24% | 66% | 56% |
| 11H30 | 65% | 84% | 76% | 69% | 78% | 83% | 75% | 72% | 84% | 52% | 78% | 66% | 23% | 66% | 56% |
| t 12H | 63% | 80% | 72% | 66% | 83% | 83% | 66% | 75% | 84% | 58% | 72% | 58% | 28% | 55% | 54% |
| o 12H30 | 58% | 78% | 62% | 65% | 81% | 81% | 52% | 80% | 80% | 62% | 74% | 57% | 33% | 51% | 51% |
| 13H | 57% | 74% | 58% | 55% | 81% | 83% | 41% | 81% | 77% | 67% | 71% | 58% | 34% | 53% | 54% |
| F 13H30 | 56% | 77% | 61% | 45% | 84% | 89% | 42% | 85% | 73% | 65% | 60% | 57% | 37% | 63% | 55% |
| r 14H | 67% | 81% | 65% | 53% | 85% | 88% | 40% | 83% | 71% | 68% | 62% | 57% | 40% | 72% | 59% |
| i 14H30 | 78% | 86% | 70% | 63% | 85% | 90% | 36% | 84% | 74% | 70% | 62% | 45% | 54% | 76% | 61% |
| d 15H | 84% | 90% | 72% | 68% | 83% | 87% | 35% | 82% | 73% | 65% | 63% | 39% | 63% | 77% | 68% |
| a 15H30 | 81% | 89% | 71% | 71% | 75% | 83% | 35% | 80% | 72% | 61% | 61% | 30% | 64% | 80% | 73% |
| y 16H | 73% | 86% | 70% | 70% | 64% | 84% | 25% | 80% | 67% | 52% | 54% | 14% | 64% | 73% | 75% |
| 16H30 | 61% | 78% | 58% | 67% | 50% | 78% | 24% | 73% | 67% | 47% | 45% | 8% | 59% | 70% | 74% |
| 8H | 85% | 85% | 98% | 62% | 82% | 98% | 93% | 86% | 63% | 46% | 68% | 76% | 64% | 77% | |
| 8H30 | 38% | 85% | 98% | 53% | 73% | 100% | 80% | 88% | 55% | 48% | 32% | 39% | 62% | 65% | 63% |
| 9H | 24% | 84% | 93% | 44% | 70% | 96% | 85% | 91% | 50% | 32% | 13% | 13% | 40% | 62% | 54% |
| 9H30 | 22% | 87% | 91% | 56% | 59% | 96% | 89% | 93% | 52% | 27% | 21% | 16% | 33% | 67% | 48% |
| 10H | 29% | 82% | 86% | 45% | 46% | 84% | 93% | 86% | 50% | 27% | 34% | 20% | 35% | 75% | 46% |
| S 10H30 | 47% | 80% | 75% | 47% | 38% | 77% | 89% | 91% | 50% | 29% | 46% | 20% | 51% | 62% | 43% |
| a 11H | 56% | 73% | 71% | 56% | 41% | 77% | 76% | 91% | 54% | 36% | 61% | 25% | 71% | 45% | 36% |
| t 11H30 | 64% | 80% | 66% | 56% | 36% | 66% | 64% | 91% | 57% | 48% | 66% | 29% | 87% | 51% | 36% |
| u 12H | 73% | 78% | 68% | 55% | 45% | 70% | 55% | 93% | 66% | 55% | 79% | 27% | 95% | 42% | 43% |
| r 12H30 | 64% | 82% | 55% | 62% | 48% | 64% | 36% | 95% | 59% | 68% | 73% | 34% | 96% | 40% | 38% |
| d 13H | 62% | 87% | 54% | 58% | 48% | 64% | 27% | 96% | 63% | 66% | 68% | 30% | 91% | 49% | 32% |
| a 13H30 | 62% | 89% | 55% | 64% | 55% | 61% | 27% | 95% | 59% | 68% | 68% | 46% | 89% | 60% | 29% |
| y 14H | 64% | 93% | 55% | 67% | 52% | 71% | 25% | 93% | 66% | 61% | 68% | 45% | 64% | 64% | 30% |
| 14H30 | 65% | 89% | 61% | 65% | 41% | 77% | 29% | 95% | 71% | 64% | 77% | 54% | 38% | 64% | 39% |
| 15H | 71% | 78% | 68% | 62% | 36% | 88% | 31% | 96% | 70% | 63% | 71% | 55% | 27% | 71% | 54% |
| 15H30 | 75% | 71% | 75% | 67% | 30% | 86% | 47% | 93% | 71% | 61% | 79% | 57% | 24% | 85% | 63% |
| 16H | 76% | 69% | 75% | 67% | 23% | 89% | 45% | 91% | 66% | 55% | 70% | 54% | 27% | 84% | 66% |
| 16H30 | 75% | 55% | 70% | 62% | 23% | 84% | 38% | 88% | 50% | 50% | 59% | 45% | 35% | 80% | 73% |

Figure 2. Heat map of time slots with absence probability during working hour (Weekdays vs Saturday)

From the figure, one can easily observe that during a day some time windows are relatively more favorable for home delivery and such time windows are different for each customer. Accordingly a suitable best delivery time window profile for every customer can be defined.

4.2 Stage 2

4.2.1 Delivery requests

This part aims to set up delivery request for the 15 customers identified in Stage 1. A request is defined by four attributes: size, origin and destination, and delivery option. In order to describe origin and destinations, we use a two-dimension (x,y) plan to simulate a city of $18*18 \text{ km}^2$, in which the 15 customers are randomly located (see Figure 3 and Table 1). In particular (0,0) represents e-tailer's storage point for the city.

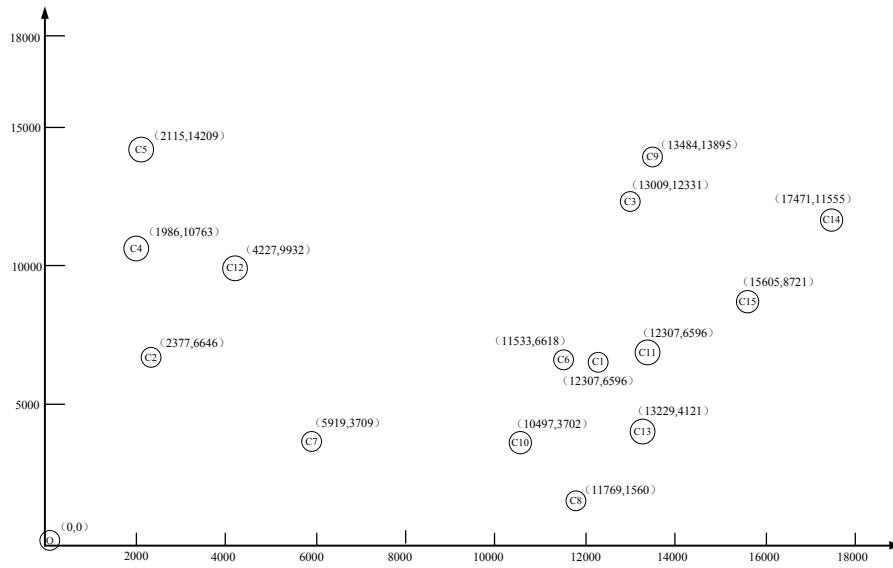


Figure 3. Geolocation of e-tailer's storage point and customers

| Customer | Order Size | Coordinate | | Optimal Time windows in Weekdays | | | Optimal Time windows in Weekdays or Saturday | | |
|----------|------------|------------|-------|----------------------------------|-------|---------------------|--|-------|---------------------|
| | | x | y | a_i | b_i | Absence probability | a_i | b_i | Absence probability |
| C1 | 1 | 12307 | 6596 | 60 | 120 | 26% | 60 | 120 | 23% |
| C2 | 1 | 2377 | 6646 | 300 | 360 | 75% | 480 | 540 | 62% |
| C3 | 1 | 13009 | 12331 | 300 | 360 | 59% | 300 | 360 | 54% |
| C4 | 1 | 1986 | 10763 | 0 | 60 | 36% | 0 | 60 | 36% |
| C5 | 1 | 2115 | 14209 | 480 | 540 | 52% | 480 | 540 | 23% |
| C6 | 1 | 11533 | 6618 | 480 | 540 | 81% | 300 | 360 | 62% |
| C7 | 1 | 5919 | 3709 | 480 | 540 | 24% | 480 | 540 | 24% |
| C8 | 1 | 11769 | 1560 | 60 | 120 | 65% | 60 | 120 | 65% |
| C9 | 1 | 13484 | 13895 | 0 | 60 | 55% | 120 | 180 | 50% |
| C10 | 1 | 10497 | 3702 | 60 | 120 | 37% | 120 | 180 | 28% |
| C11 | 1 | 13321 | 6989 | 0 | 60 | 40% | 60 | 120 | 17% |
| C12 | 1 | 4227 | 9932 | 480 | 540 | 11% | 480 | 540 | 11% |
| C13 | 1 | 13229 | 4121 | 180 | 240 | 23% | 180 | 240 | 23% |
| C14 | 1 | 17471 | 11555 | 0 | 60 | 45% | 240 | 300 | 41% |
| C15 | 1 | 15605 | 8721 | 240 | 300 | 52% | 300 | 360 | 30% |

Table 1. Input data

The objective of this stage is to establish delivery routes that satisfy all delivery requests, while minimizing total distance generated by the first-round delivery and, in case of failure, the rescheduled delivery. Several assumptions are made here.

- (1) Every customer has one delivery request of size 1 in a week and the truck capacity is 5.
- (2) The delivery option given to the customers would be the day of delivery expected. Then, it is up to the carrier to select the optimal delivery time on that day and propose it to the customers. Two types of days are considered

here: weekdays and Saturday. By this assumption we attempt to explore the interest of Saturday delivery.

- (3) All customers will accept the proposed delivery time windows. The time window is set to 1 hour and the associated absence probability is the average of the values in Figure 2, e.g. the average value of 8h-8h30 and 8h30-9h for 8h-9h.
- (4) Failed delivery due to the absence of customer will be rescheduled as a direct delivery from the e-tailer's storage point to the customer in the next day. Accordingly, the distance generated by the rescheduling is equal to *round trip distance * absence probability* for each customer. In other words, we do not consider another VRPTW for the failed deliveries due to the service constraint and the lack of knowledge on new deliveries.
- (5) Truck's speed is set to 20 km/h in city.
- (6) Service time per customer is set to 5 minutes.

4.2.2 Delivery optimization model

We assume that in each time slot the delivery customers' attendance probability is known and that the optimal time slot for each customer's delivery is that with the highest attendance probability. It is therefore a typical capacitated VRPTW (Baldacci et al., 2011, Solomon, 1987, Azi et al., 2007), in which the time window is the optimal time slot for delivery. We propose a Mixed Integer Linear Programming (MILP) model for the capacitated VRPTW which follows the guidelines in Toth and Vigo (2002) and incorporates the classical constraints to enforce time windows that can be found in Azi et al. (2007). The MILP can be described as follows. Given a set of customers $V=\{1,2,...,n\}$ with known demands of q_i for any $i \in V$, we have a fleet of homogeneous vehicles of capacity Q to deliver those demands, from a depot noted as 0 to customers. The directed graph can be thus noted as $G=(V^+,A)$, where $V^+=V \cup \{0\}$ is the set of nodes and A is the set of arcs. Each arc $(i,j) \in A$ is associated with a travel time $t_{ij} > 0$ and a distance $d_{ij} > 0$. Each customer $i \in V$ is associated with a service time s_i and a time window $[a_i, b_i]$ that presents respectively the earliest and latest time at which the service must begin at i . The objective is to minimize the total distance traveled to serve all customers while satisfying the capacity and time window constraints

$$\begin{aligned}
& \text{Min } \sum_{i \in V^+} \sum_{j \in V^+} d_{ij} x_{ij} & (3) \\
\text{s.t. } & \sum_{j \in V^+} x_{ij} = 1, \quad i \in V & (4) \\
& \sum_{i \in V^+} x_{ih} - \sum_{j \in V^+} x_{hj} = 0, \quad h \in V & (5) \\
& q_i \leq u_i \leq Q, \quad i \in V & (6) \\
& u_i - u_j + Qx_{ij} \leq Q - q_j, \quad i, j \in V & (7) \\
& t_i + (t_{ij} + s_i)x_{ij} - N(1 - x_{ij}) \leq t_j, \quad i, j \in V & (8) \\
& a_i \leq t_i \leq b_i, \quad i \in V & (9) \\
& x_{ij} \in \{0,1\}, \quad i, j \in V^+ & (10) \\
& u_i, t_i \geq 0, \quad i \in V & (11)
\end{aligned}$$

With decision variables: x_{ij} is 1 if arc (i,j) is included in any route, 0 otherwise; t_i indicates the start time of service on every node i .

In the model, (4)-(5) guarantee that every customer is visited exactly once and that every route begins from the depot and ends at the depot. Eq (6)-(7) ensure that the total demand on every route will not exceed the vehicle capacity Q , in particular u_i is a variable indicating the accumulative total demand on customer i . Eq (8)-(9) ensure feasibility of the time schedule on every route, with N a large number.

4.2.3 Results and optimal transportation plan

Five scenarios are designed for this case, as shown in Table 2. S0 serves as baseline scenario that does not consider time windows in transportation planning (thus not taking into account absence probabilities). In other words, S0 is a classic VRP optimizing transportation distance and it represents a conventional approach to the problem. S1 and S2 take into account the customers' optimal time windows profile only in weekdays. S3 and S4 consider both weekdays and Saturday so that customers are divided in two clusters: one for the customers with highest attendance probability appeared on weekdays and another one for the others. Scenarios 1-4 are therefore used as different ways the proposed approach can be implemented in this business scenario.

All time windows can be observed from Figure 2. Since the optimal time windows can be dispersive in a day, in S2 and S4 we deliberately added the constraint (12) to limit the waiting time in-between two successive customers to 60 minutes. Constraints (8) and (12) define that, when $x_{ij}=1$, $t_j - 60 \leq t_i + s_i + t_{ij} \leq t_j$.

$$t_i + (t_{ij} + s_i)x_{ij} - N(1 - x_{ij}) \geq t_j - 60, \quad i, j \in V \quad (12)$$

All scenarios ran using GUSEK on a ThinkPad T440 with 4 GB of RAM. For all scenarios, every single computation process required about one minute, except S0 that required nearly 7 hours. The results are summarized in Table 2 and Table 3.

| Sc. | Time Windows | Saturday Delivery | Waiting Time ≤ 60 mn | Distance of First Delivery (km) | Total Distance (km) | Δ km vs S0 | Average Probability of Successful First Delivery |
|-----|--------------|-------------------|---------------------------|---------------------------------|---------------------|-------------------|--|
| S0 | N | N | N | 111 | 367 | - | 37% |
| S1 | Y | N | N | 130 | 320 | -13% | 55% |
| S2 | Y | N | Y | 166 | 356 | -3% | 55% |
| S3 | Y | Y | N | 143 | 295 | -20% | 63% |
| S4 | Y | Y | Y | 176 | 327 | -11% | 63% |

Table 2. Scenarios and results

| Sc. | Optimal Tours |
|-----|--|
| S0 | R1=(0-10-6-1-13-8-0), R2=(0-7-12-5-4-2-0), R3=(0-11-15-14-9-3-0) |
| S1 | R1=(0-4-2-5-12-7-0), R2=(0-8-10-1-13-6-0), R3=(0-11-9-14-15-3-0) |
| S2 | R1=(0-7-0); R2=(0-4-0), R3=(0-6-12-5-0), R4=(0-8-13-15-3-2-0), R5=(0-9-14-11-1-10-0) |
| S3 | WD: R1=(0-4-8-13-12-7-0); Saturday: R2=(0-1-11-14-15-6-0), R3=(0-10-9-3-5-2-0) |
| S4 | WD: R1=(0-4-8-13-0), R2=(0-12-7-0); Saturday: R3=(0-5-2-0), R4=(0-1-11-10-0), R5=(0-9-14-3-15-6-0) |

Table 3. Optimal tours in the scenarios

Theoretically S0 provides the shortest routes without considering time windows and failure probability, i.e. the first-round delivery. However, this routing plan caused a low rate of successful delivery of 37% and, as a result, generated the highest total distance resulted from a significant number of rescheduled deliveries. Low rate of successful delivery can also be seen as poor service to customers, since it means that customers might not receive their orders on time. In terms of total distance, S3 considering Saturday delivery performs the best, reducing 20% km comparing to S0, thanks to a higher rate of successful delivery of 63%. This is because 9 of the 15 customers have lower absence probability on Saturday. As shown by S2 vs S1, or S4 vs S3, the constraint (12) that aims to limit each waiting time made total distance increased. However, without this constraint, the waiting time in-between two successive customers can be longer than 3 hours in S1 and S3 (result of timeline is not provided here due to lack of space). In practice, waiting time depends closely on the quantity of customers to deliver in a tour.

5 Conclusions and Discussion

With the emergence of e-commerce in grocery retail, the food supply chain faces new challenges. In this paper, we focused on such a challenge regarding the successful delivery of grocery orders placed online. Due to the perishability and sensitivity of some grocery items, customer attendance is often critical for the successful delivery

of online orders. As a solution to this problem, this paper introduced a two-stage methodological approach that utilizes customer-related data to schedule transportation plans. This is done by first estimating the probabilities of customer attendance/absence at different point of times during a day and then using these estimations in a way that satisfies a company's key performance indicators (e.g. maximize the probability of attended delivery while minimizing travel distance covered by delivery trucks).

As an application example, this paper presented an experimental study to investigate how a customer's historical electricity consumption data can be used to estimate time windows with a lower probability of inoccupation (absence). The best time windows were then used in a VRP model in order to plan the deliveries of online orders to customers aiming at improving delivery success rate. A numerical study has been conducted to demonstrate the effectiveness of the proposed approach that shows its potential in the delivery of online grocery shopping.

Besides increasing the rate of successful deliveries, the proposed approach can help e-tailers better understand the habits of their customers and thus the optimal delivery time for them. It can also be considered as a useful tool for dynamically pricing different delivery options and as a mechanism for time slot management. From a customer's point of view, the approach can improve customer satisfaction as it can reduce unnecessary traveling to pick up missed orders or long telephone calls required to re-arrange deliveries. It is also obvious that the approach can easily be used in different business cases in urban freight transportation and last mile logistics (e.g. non-food items, general merchandise) where attended home delivery is critical and alternative solutions cannot be easily offered.

This study is among the first ones integrating data mining techniques in urban freight transportation. Some prospects can thus be identified to this line of research. For example, one may test the approach with other customer-related datasets, e.g. water consumption, historical deliveries, or with different data mining techniques in order to compare accuracy and performance. The attendance probabilities can also be used in different VRP models or for different reasons such as slot pricing. Further, the proposed methodological approach can also be generalized from e-grocery commerce

to all businesses that provide home delivery service, i.e. the general attended home delivery problem (AHDP).

Several limitations have to be carefully considered with regards to the proposed approach. Firstly, some legal issues, (e.g. data privacy and security), can arise from accessing and using energy consumption data of households by third parties. Secondly, some e-grocery retailers offer the possibility of time slot selection from the customer during the placement of an order. Even though the proposed approach is generic enough to cover this case, the application example in this paper did not demonstrate this interesting scenario. Finally, as any other data-driven approach, our approach might be limited by computational capacity. The first stage of our approach could be limited by the size of the training data and the data mining algorithms used on them. The second stage is mostly limited by algorithmic efficiency. Further field testing to select the appropriate, case-specific training data and algorithms is required before putting the approach into practice.

References

- ABREU, J. & PEREIRA, F. Household electricity consumption routines and tailored feedback. ACEEE Summer Study on Energy Efficiency in Buildings, 2012. 3-6.
- AGATZ, N., CAMPBELL, A., FLEISCHMANN, M. & SAVELSBERGH, M. 2011. Time Slot Management in Attended Home Delivery. *Transportation Science*, 45, 435-449.
- AGATZ, N. A. H., FLEISCHMANN, M. & VAN NUNEN, J. A. E. E. 2008. E-fulfillment and multi-channel distribution – A review. *European Journal of Operational Research*, 187, 339-356.
- ALLEN, J., BROWNE, M., WOODBURN, A. & LEONARDI, J. A review of urban consolidation centres in the supply chain based on a case study approach. *Supply Chain Forum: an international journal*, 2014. Taylor & Francis, 100-112.
- AMAN, S., SIMMHAN, Y. & PRASANNA, V. K. Improving energy use forecast for campus micro-grids using indirect indicators. *Data Mining Workshops (ICDMW)*, 2011 IEEE 11th International Conference on, 2011. IEEE, 389-397.
- AMAZON. 2016. *Amazon Prime Air* [Online]. Available: <https://www.amazon.com/Amazon-Prime-Air/b?ie=UTF8&node=8037720011> [Accessed].
- ANCKAR, B., WALDEN, P. & JELASSI, T. 2002. Creating customer value in online grocery shopping. *International Journal of Retail & Distribution Management*, 30, 211-220.
- ANU, B., MERVI, L. & MARKKU, T. 2012. E-Commerce Logistics: A Literature Research Review and Topics for Future Research. *International Journal of E-Services and Mobile Applications (IJESMA)*, 4, 1-22.
- AZI, N., GENDREAU, M. & POTVIN, J.-Y. 2007. An exact algorithm for a single-vehicle routing problem with time windows and multiple routes. *European journal of operational research*, 178, 755-766.
- BALDACCI, R., MINGOZZI, A. & ROBERTI, R. 2011. New Route Relaxation and Pricing Strategies for the Vehicle Routing Problem. *Operations Research*, 59, 1269-1283.
- BECKEL, C., SADAMORI, L. & SANTINI, S. Towards automatic classification of private households using electricity consumption data. *Proceedings of the Fourth ACM*

- Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings, 2012. ACM, 169-176.
- BOYER, K. K. & HULT, G. T. M. 2005. Customer Behavior in an Online Ordering Application: A Decision Scoring Model*. *Decision Sciences*, 36, 569-598.
- BOYER, K. K. & HULT, G. T. M. 2006. Customer behavioral intentions for online purchases: An examination of fulfillment method and customer experience level. *Journal of Operations Management*, 24, 124-147.
- BREUGELMANS, E. & CAMPO, K. 2016. Cross-Channel Effects of Price Promotions: An Empirical Analysis of the Multi-Channel Grocery Retail Sector. *Journal of Retailing*, 92, 333-351.
- CAMPBELL, A. M. & SAVELSBERGH, M. 2006. Incentive Schemes for Attended Home Delivery Services. *Transportation Science*, 40, 327-341.
- CATTARUZZA, D., ABSI, N., FEILLET, D. & GONZÁLEZ-FELIU, J. 2015. Vehicle routing problems for city logistics. *EURO Journal on Transportation and Logistics*, 1-29.
- CHEN, C., PAN, S., WANG, Z. & ZHONG, R. Y. 2016. Using taxis to collect citywide E-commerce reverse flows: a crowdsourcing solution. *International Journal of Production Research*, 1-12.
- COLLA, E. & LAPOULE, P. 2012. E-commerce: exploring the critical success factors. *International Journal of Retail & Distribution Management*, 40, 842-864.
- DE MARCO, A., CAGLIANO, A. C., MANGANO, G. & PERFETTI, F. 2014. Factor Influencing Logistics Service Providers Efficiency' in Urban Distribution Systems. *Transportation Research Procedia*, 3, 499-507.
- DE SOUZA, R., GOH, M., LAU, H.-C., NG, W.-S. & TAN, P.-S. 2014. Collaborative urban logistics—synchronizing the last mile a Singapore research perspective. *Procedia-Social and Behavioral Sciences*, 125, 422-431.
- DHL. 2016. *DHL Parcelcopter 3.0* [Online]. Available: http://www.dpdhl.com/en/media_relations/specials/parcelcopter.html [Accessed].
- EHMKE, J. F. 2012a. Attended Home Delivery. *Integration of Information and Optimization Models for Routing in City Logistics*. Springer US.
- EHMKE, J. F. 2012b. *Integration of information and optimization models for routing in city logistics*, Springer Science & Business Media.
- EHMKE, J. F. & CAMPBELL, A. M. 2014. Customer acceptance mechanisms for home deliveries in metropolitan areas. *European Journal of Operational Research*, 233, 193-207.
- EHMKE, J. F. & MATTFELD, D. C. 2012. Vehicle Routing for Attended Home Delivery in City Logistics. *Procedia - Social and Behavioral Sciences*, 39, 622-632.
- EHMKE, J. F., MEISEL, S. & MATTFELD, D. C. 2012. Floating car based travel times for city logistics. *Transportation Research Part C: Emerging Technologies*, 21, 338-352.
- FERNIE, J., SPARKS, L. & MCKINNON, A. C. 2010. Retail logistics in the UK: past, present and future. *International Journal of Retail & Distribution Management*, 38, 894-914.
- GEVAERS, R., VAN DE VOORDE, E. & VANELSLANDER, T. 2011. Characteristics and typology of last-mile logistics from an innovation perspective in an urban context. *City Distribution and Urban Freight Transport: Multiple Perspectives*, Edward Elgar Publishing, 56-71.
- HSU, C.-I., HUNG, S.-F. & LI, H.-C. 2007. Vehicle routing problem with time-windows for perishable food delivery. *Journal of Food Engineering*, 80, 465-475.
- HÜBNER, A., KUHN, H. & WOLLENBURG, J. 2016. Last mile fulfilment and distribution in omni-channel grocery retailing: A strategic planning framework. *International Journal of Retail & Distribution Management*, 44, 228-247.
- ISHFAQ, R., DEFEE, C. C., GIBSON, B. J. & RAJA, U. 2016. Realignment of the physical distribution process in omni-channel fulfillment. *International Journal of Physical Distribution & Logistics Management*, 46, 543-561.

- IWAN, S., KIJEWSKA, K. & LEMKE, J. 2016. Analysis of Parcel Lockers' Efficiency as the Last Mile Delivery Solution – The Results of the Research in Poland. *Transportation Research Procedia*, 12, 644-655.
- KANG, C., MOON, J., KIM, T. & CHOE, Y. Why Consumers Go to Online Grocery: Comparing Vegetables with Grains. 2016 49th Hawaii International Conference on System Sciences (HICSS), 5-8 Jan. 2016 2016. 3604-3613.
- KARAKOSTAS, B., KARDARAS, D. & PAPATHANASSIOU, E. 2005. The state of CRM adoption by the financial services in the UK: an empirical investigation. *Information & Management*, 42, 853-863.
- KLEIMINGER, W., BECKEL, C., STAAKE, T. & SANTINI, S. 2013. Occupancy Detection from Electricity Consumption Data. *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings*. Roma, Italy: ACM.
- KOLTER, J. Z. & FERREIRA JR, J. 2011. A large-scale study on predicting and contextualizing building energy usage.
- KOSTER, R. B. M. D. 2003. Distribution strategies for online retailers. *IEEE Transactions on Engineering Management*, 50, 448-457.
- LEE, H. L. & WHANG, S. 2001. Winning the last mile of e-commerce. *MIT Sloan Management Review*, 42, 54.
- LIAO, S.-H., CHEN, Y.-J. & LIN, Y.-T. 2011. Mining customer knowledge to implement online shopping and home delivery for hypermarkets. *Expert Systems with Applications*, 38, 3982-3991.
- LOWE, R. & RIGBY, M. 2014. The Last Mile: Exploring the online purchasing and delivery journey. UK.
- MITHAS, S., KRISHNAN, M. S. & FORNELL, C. 2005. Why do customer relationship management applications affect customer satisfaction? *Journal of Marketing*, 69, 201-209.
- MORTIMER, G., FAZAL E HASAN, S., ANDREWS, L. & MARTIN, J. 2016. Online grocery shopping: the impact of shopping frequency on perceived risk. *The International Review of Retail, Distribution and Consumer Research*, 26, 202-223.
- OGAWARA, S., CHEN, J. C. & ZHANG, Q. 2003. Internet grocery business in Japan: current business models and future trends. *Industrial Management & Data Systems*, 103, 727-735.
- PALOHEIMO, H., LETTENMEIER, M. & WARIS, H. 2015. Transport reduction by crowdsourced deliveries – a library case in Finland. *Journal of Cleaner Production*.
- PUNAKIVI, M. & SARANEN, J. 2001. Identifying the success factors in e-grocery home delivery. *International Journal of Retail & Distribution Management*, 29, 156-163.
- RAMANATHAN, R. 2010. The moderating roles of risk and efficiency on the relationship between logistics performance and customer loyalty in e-commerce. *Transportation Research Part E: Logistics and Transportation Review*, 46, 950-962.
- RICKER, F. & KALAKOTA, R. 1999. Order fulfillment: the hidden key to e-commerce success. *Supply Chain Management Review*, 11, 60-70.
- SASKIA, S., MAREĬ, N. & BLANQUART, C. 2016. Innovations in e-grocery and Logistics Solutions for Cities. *Transportation Research Procedia*, 12, 825-835.
- SCHÖLKOPF, B., PLATT, J. C., SHAW-TAYLOR, J., SMOLA, A. J. & WILLIAMSON, R. C. 2001. Estimating the support of a high-dimensional distribution. *Neural computation*, 13, 1443-1471.
- SOLOMON, M. M. 1987. Algorithms for the vehicle routing and scheduling problems with time window constraints. *Operations research*, 35, 254-265.
- TANSKANEN, K., YRJÖLÄ, H. & HOLMSTRÖM, J. 2002. The way to profitable Internet grocery retailing – six lessons learned. *International Journal of Retail & Distribution Management*, 30, 169-178.
- TOTH, P. & VIGO, D. 2002. Models, relaxations and exact approaches for the capacitated vehicle routing problem. *Discrete Applied Mathematics*, 123, 487-512.

- VAN DUIN, J., QUAK, H. & MUÑUZURI, J. 2010. New challenges for urban consolidation centres: A case study in The Hague. *Procedia-Social and Behavioral Sciences*, 2, 6177-6188.
- WANG, C., MAO, Z., O'KANE, J. & WANG, J. 2016a. An exploration on e-retailers' home delivery – strategic elements and their prioritisation. *Business Process Management Journal*, 22, 614-633.
- WANG, G., GUNASEKARAN, A., NGAI, E. W. T. & PAPADOPOULOS, T. 2016b. Big data analytics in logistics and supply chain management: Certain investigations for research and applications. *International Journal of Production Economics*, 176, 98-110.
- WOOD, G. & NEWBOROUGH, M. 2003. Dynamic energy-consumption indicators for domestic appliances: environment, behaviour and design. *Energy and Buildings*, 35, 821-841.
- ZHANG, Y., FANG, Y., WEI, K.-K., RAMSEY, E., MCCOLE, P. & CHEN, H. 2011. Repurchase intention in B2C e-commerce—A relationship quality perspective. *Information & Management*, 48, 192-200.
- ZOTT, C., AMIT, R. & DONLEVY, J. 2000. Strategies for value creation in e-commerce:: best practice in Europe. *European Management Journal*, 18, 463-475.